APPLICATION

FOR

UNITED STATES LETTERS PATENT

TITLE: DELIVERY OF STREAMING MEDIA

APPLICANT: SUDHEER SIRIVARA, JEFFREY MCVEIGH, BOB J. REESE, GIANNI G. FERRISE, PHILLIP G. AUSTIN, RAM R. RAO AND SHOBHANA SUBRAMANIAN

CERTIFICATE OF MAILING BY EXPRESS MAIL

Express Mail Label No. EL224699429US

I hereby certify under 37 CFR §1.10 that this correspondence is being deposited with the United States Postal Service as Express Mail Post Office to Addressee with sufficient postage on the date indicated below and is addressed to the Commissioner for Patents, Washington, D.C. 20231.

May 30, 2001

Date of Deposit

Signature

Joshua Cronin

Typed or Printed Name of Person Signing Certificate

Atty. Docket No: 10559-477001
Client Docket No: P-11156

DELIVERY OF STREAMING MEDIA

FIELD OF INVENTION

This invention relates to delivery of streaming media.

BACKGROUND

Streaming media refers to content, typically audio, video, or both, that is intended to be displayed to an end-user as it is transmitted from a content provider. Because the content is being viewed in real-time, it is important that a continuous and uninterrupted stream be provided to the user. The extent to which a user perceives an uninterrupted stream that displays uncorrupted media is referred to as the "Quality of Service", or QOS, of the system.

A content delivery service typically evaluates its QOS by collecting network statistics and inferring, on the basis of those network statistics, the user's perception of a media stream. These network statistics include such quantities as packet loss and latency that are independent on the nature of the content. The resulting evaluation of QOS is thus content-independent.

BRIEF DESCRIPTION OF THE FIGURES

FIGS. 1 and 2 show content delivery systems.

DETAILED DESCRIPTION

As shown in FIG. 1, a content delivery system **10** for the

delivery of a media stream **12** from a content server **14** to a client **16** includes two distinct processes. Because a media stream requires far more bandwidth than can reasonably be accommodated on today's networks, it is first passed through an encoder **18** executing on the content server **14**. The encoder **18** transforms the media stream **12** into a compressed form suitable for real-time transmission across a global computer network **22**. The resulting encoded media stream **20** then traverses the global computer network **22** until it reaches the client **16**. Finally, a decoder **24** executing on the client **16** transforms the encoded media stream **20** into a decoded media stream **26** suitable for display.

In the content delivery system **10** of FIG. 1, there are at least two mechanisms that can impair the media stream. First, the encoder **18** and decoder **24** can introduce errors. For example, many encoding processes discard high-frequency components of an image in an effort to compress the media stream **12**. As a result, the decoded media stream **26** may not be a replica of the original media stream **12**. Second, the vagaries of network transmission, many of which are merely inconvenient when text or static images are delivered, can seriously impair the real-time delivery of streaming media.

These two impairment mechanisms, hereafter referred to as encoding error and transmission error, combine to affect the end-user's subjective experience in viewing streaming media. However,

the end-user's subjective experience also depends on one other factor thus far not considered: the content of the media stream **12** itself.

The extent to which a particular error affects an end-user's enjoyment of a decoded media stream **26** depends on certain features of the media stream **12**. For example, a media stream **12** rich in detail will suffer considerably from loss of sharpness that results from discarding too many high frequency components. In contrast, the same loss of sharpness in a media stream **12** rich in impressionist landscapes will scarcely be noticeable.

Referring to FIG. 2, a system **28** incorporating the invention includes a content-delivery server **30** in data communication with a client **32** across a global computer network **34**. The system **28** also includes an aggregating server **36** in data communication with both the client **32** and the content-delivery server **30**. The link between the aggregating server **36** and the client **32** is across the global computer network **34**, whereas the link between the aggregating server **36** and the content-delivery server **30** is typically over a local area network.

An encoder **38** executing on the content-delivery server **30** applies an encoding or compression algorithm to the original media stream **39**, thereby generating an encoded media stream **40**. For simplicity, FIG. 2 is drawn with the output of the encoder **38**

leading directly to the global computer network **34**, as if encoding occurred in real-time. Although it is possible, and sometimes desirable, to encode streaming media in real-time (for example in the case of video-conferencing applications), in most cases encoding is carried out in advance. In such cases, the encoded media **40** is stored on a mass-storage system (not shown) associated with the content-delivery server **30**.

A variety of encoding processes are available. In many cases, these encoding processes are lossy. For example, certain encoding processes will discard high-frequency components of an image under the assumption that, when the image is later decoded, the absence of those high-frequency components will not be apparent to the user. Whether this is indeed the case will depend in part on the features of the image.

In addition to being transmitted to the client **32** over the global computer network **34**, the encoded media **40** at the output of the encoder **38** is also provided to the input of a first decoder **42**, shown in FIG. 2 as being associated with the aggregating server **36**. The first decoder **42** recovers the original media stream to the extent that the possibly lossy encoding performed by the encoder **38** makes it possible to do so.

The output of the decoding process is then provided to a first feature extractor **44**, also executing on the aggregating server **36**.

The first feature extractor **44** implements known feature extraction algorithms for extracting temporal or spatial features of the encoded media **40**. Known feature extraction methods include the Sarnoff JND ("Just Noticeable Difference") method and the methods disclosed in ANSI T1.801.03-1996 ("American National Standard for Telecommunications - Digital Transport of One Way Video Signals - Parameters for Objective Performance Specification") specification.

A typical feature-extractor might evaluate a discrete cosine transform ("DCT") of an image or a portion of an image. The distribution of high and low frequencies in the DCT would provide an indication of how much detail is in any particular image. Changes in the distribution of high and low frequencies in DCTs of different images would provide an indication of how rapidly images are changing with time, and hence how much "action" is actually in the moving image.

The original media **39** is also passed through a second feature extractor **46** identical to the first feature extractor **44**. The outputs of the first and second feature extractors **44, 46** are then compared by a first analyzer **48**. This comparison results in the calculation of an encoding metric indicative of the extent to which the subjective perception of a user would be degraded by the encoding and decoding algorithms by themselves.

An analyzer compares DCTs of two images, both of which are

typically matrix quantities, and maps the difference to a scalar. The output of the analyzer is typically a dimensionless quantity between 0 and 1 that represents a normalized measure of how different the frequency distribution of two images are.

The content-delivery server **30** transmits the encoded media **40** to the user by placing it on the global computer network **34.** Once on the global computer network **34,** the encoded media **40** is subjected to the various difficulties that are commonly encountered when transmitting data of any type on such a network **34.** These include jitter, packet loss, and packet latency. In one embodiment, statistics on these and other measures of transmission error are collected by a network performance monitor **52** and made available to the aggregating server **36.**

The media stream received by the client **32** is then provided to a second decoder **54** identical to the first decoder **42.** A decoded stream **56** from the output of the second decoder **54** is made available for display to the end-user. In addition, the decoded stream **56** is passed through a third feature extractor **58** identical to the first and second feature extractors **44, 46.** The output of the third feature extractor **58** is provided to a second analyzer **60.**

The inputs to both the first and third feature extractor **44, 58** have been processed by the same encoder **38** and by identical decoders **42, 54.** However, unlike the input to the third feature

extractor **58**, the input to the first feature extractor **44** was never transported across the network **34**. Hence, any difference between the outputs of the first and third feature extractors **44, 58** can be attributed to transmission errors alone. This difference is determined by second analyzer **60**, which compares the outputs of the first and third feature extractors **44, 58**. On the basis of this difference, the second analyzer **60** calculates a transmission metric indicative of the extent to which the subjective perception of a user would be degraded by the transmission error alone.

The system **28** thus provides an estimate of a user's perception of the quality of a media stream on the basis of features in the rendered stream. This estimate is separable into a first portion that depends only on encoding error and a second portion that depends only on transmission error.

Having determined a transmission metric, it is useful to identify the relative effects of different types of transmission errors on the transmission metric. To do so, the network statistics obtained by the network performance monitor **52** and the transmission metric determined by the second analyzer **60** are provided to a correlator **62**. The correlator **62** can then correlate the network statistics with values of the transmission metric. The result of this correlation identifies those types of network errors that most significantly affect the end-user's experience.

In one embodiment, the correlator **62** averages network statistics over a fixed time-interval and compares averages thus generated with corresponding averages of transmission metrics for that time-interval. This enables the correlator **62** to establish, for that time interval, contributions of specific network impairments, such as jitter, packet loss, and packet latency, toward the end-user's experience.

Although the various processes are shown in FIG. 1 as executing on specific servers, this is not a requirement. For example, the system **28** can also be configured so that the first decoder **42** executes on the content-delivery server **30** rather than on the aggregating server **36** as shown in FIG. 1. In one embodiment, the output of the first feature extractor is sent to the client and the second analyzer executes at the client rather than at the aggregating server **36**. The server selected to execute a particular process depends, to a great extent, on load balancing.

Other embodiments are within the scope of the following claims.

We claim: